DOCUMENT RESUME

ED 348 339 SP 033 897

AUTHOR Tyson, LeaAnn; Silverman, Stephen
TITLE Differences in Texas Teacher Appraisal System Scores.
PUB DATE 21 Apr 92
NOTE 48p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, CA, April 20-24, 1992).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Analysis of Variance; Career Ladders; Classroom Observation Techniques; Elementary School Teachers; Elementary Secondary Education; Interrater Reliability; *Scores; Secondary School Teachers; Teacher Effectiveness; *Teacher Evaluation; *Testing Problems; *Test Reliability; *Test Validity
IDENTIFIERS *Texas Teacher Appraisal System

ABSTRACT

The purpose of this study was to investigate differences in Texas Teacher Appraisal System (TTAS) scores when considering the scores of the first four individual domains (Instructional Strategies, Management and Organization, Presentation of Subject Matter, and Learning Environment), the sum of the scores of Domains I through IV, and the overall summary performance score for teachers in a large school district over a period of 2 years. Specifically, scores between appraiser types (primary or second), levels (elementary or secondary), years (1988-1989 or 1989-1990), and apraisal periods (first or second) were investigated. The subjects were teachers in a large school district in Texas for the school years 1988-1989 (N=1,159) and 1989-1990 (N=1,207). All teachers were evaluated with the TTAS by a certified appraiser, following the procedures mandated by the State Board of Education. Campus administrators or central office personnel conducted all observations. MANOVA and ANOVA were performed to examine differences in the scores of various subgroups of teachers. In general, results showed that significantly higher scores were awarded by the primary appraisers and that elementary teachers received higher scores than secondary teachers. In addition, scores increased from the first appraisal period to the second appraisal period and from year to year. The results of this study suggest the need for additional research on the validity of the instrument, staff development for teachers and appraisers, and modifications to the instrument or scoring procedures to make TTAS scores meaningful. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED348339

DIFFERENCES IN TEXAS TEACHER APPRAISAL SYSTEM SCORES[1]

LeaAnn Tyson
Western Washington University
Bellingham, WA

Stephen Silverman
University of Illinois at Urbana-Champaign
Urbana, IL

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Paper presented at the annual meeting of the American Educational Research Association, April 21, 1992, San Francisco.

[1] This research was conducted at The University of Texas at Austin for the doctoral dissertation by the first author under the direction of the second author.

SP033697

BEST COPY AVAILABLE

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
LeaAnn Tyson
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

ABSTRACT

The purpose of this study was to investigate differences in Texas Teacher Appraisal System (TTAS) scores when considering the scores of the first four individual domains (Instructional Strategies, Management and Organization, Presentation of Subject Matter, and Learning Environment), the sum of the scores of Domains I through IV, and the overall summary performance score for teachers in a large school district over a period of two years. Specifically, scores between appraiser types (primary or second), levels (elementary or secondary), years (1988-1989 or 1989-1990) and appraisal periods (first or second) were investigated. The subjects were teachers in a large school district in Texas for the school years 1988-1989 (N=1159) and 1989-1990 (N=1207). All teachers were evaluated with the TTAS by a certified appraiser, following the procedure mandated by the State Board of Education. Campus administrators or central office personnel conducted all observations. MANOVA and ANOVA were performed to examine differences in the scores of various subgroups of teachers. In general, results showed significantly higher scores were awarded by the primary appraisers, and elementary teachers received higher scores than secondary teachers. In addition, scores increased from the first appraisal period to the second appraisal period and from year to year. The results of this study suggest additional research on the validity of the instrument, staff development for teachers and appraisers, and modifications to the instrument or scoring procedures to make TTAS scores meaningful.

DIFFERENCES IN TEXAS TEACHER APPRAISAL SYSTEM SCORES

For administrators in any profession, evaluating the effectiveness of personnel is a difficult, yet important, task. In education, teacher assessment has received increased attention from the media, policy makers, and from practitioners themselves in the past ten to fifteen years (Ellett, 1990). This greater concern and interest in the evaluation of teachers' work has resulted from accountability issues that focus on the credibility and quality of teachers. Education policy makers consistently link better teachers to better education. Therefore, it is assumed that a key to improving education is to upgrade the quality of teachers.

There are few issues that teachers, policy makers, administrators and parents agree upon regarding the evaluation of teachers. Should the assessment be the same for all teachers? What data should be collected? How often should evaluations occur? How should the process take place? How should performances be judged? Everyone seems to agree, however, that evaluation should take place and the evaluators should be trained and skilled.

The major purpose of teacher evaluation is to improve instruction, although this assessment also generates data for administrative decisions such as teacher dismissal, contract renewal, and more recently, determining merit pay and career ladder placements. Evaluating the effectiveness of teachers may involve many aspects of assessment such as competency testing, student achievement measures, peer evaluation, self-assessment, and examination of teacher-produced documents. Many suggest that evaluation should not be based on a single assessment, but most agree that teacher observation is critical in evaluation. Consequently, observation of

teaching is by far the most common method of teacher assessment (Ellett & Garland, 1986; Wise, Darling-Hammond, McLaughlin & Bernstein, 1984).

The period of the late 1950's through the 1970's produced numerous classroom observation instruments as a result of research on classroom teaching (Brophy & Good, 1986; Medley & Mitzel, 1963; Rosenshine & Furst, 1973). Because of the availability of instruments, empirical research findings, and demands for accountability through demonstrated teacher competency, legislatively-enacted development of large-scale teacher evaluation systems followed in the southern states of South Carolina, Florida, North Carolina, Tennessee, Mississippi, Kentucky and Virginia (Ellett, 1990). As many as ten states have implemented large-scale assessment procedures utilizing observation instruments to measure teaching performance, while several other states are in the developmental phase (Association of Teacher Educators, 1988). In addition, several states utilize teacher observation assessment instruments for teachers seeking initial certification (United States Department of Education, 1987). This method of evaluation has evolved to focus on teacher competencies and their mastery (Ellett, 1990).

To date, the largest state to implement statewide teacher appraisal using observation is Texas. The Texas Teacher Appraisal System (TTAS) was implemented statewide in 1986 and is based, for the most part, on classroom teaching performance (Texas Education Agency, 1988). The TTAS contains 5 domains, broken down into 13 criteria, which are further divided into 65 behavioral indicators (Texas Education Agency, 1988). The first four domains (Instructional Strategies, Classroom Management and Organization, Presentation of Subject Matter,

and Learning Environment) are based on classroom teaching performance, and the fifth domain is related to professional growth.

The Texas Career Ladder, a merit pay system of three levels, is directly related to the TTAS. Although one program could exist without the other, the TTAS and the career ladder are often viewed as one because the performance score on the TTAS is one criteria for advancement and maintenance at the three career ladder levels. Years of teaching experience and hours of professional growth serve as the other two criteria (Texas Education Agency, 1988).

One assumption of the Texas teacher appraisal process is that trained observers (appraisers) can observe a teacher independently and arrive at a similar score. Concerns about the reliability of observations are frequent among administrators, teachers, policy makers, and teacher organizations, particularly when the observation data is utilized in a summative manner for administrative decisions regarding contract renewal and career ladder criteria. For the school year 1990-1991, this concern has resulted in a "third appraiser rule" which requires districts to provide a third observation of a teacher if there is at least a fifteen point difference in the scores of two appraisers on the first four of the five TTAS domains (Texas Education Code sec. 13.302-13.321, 1990). This suggests that one appraiser failed to accurately evaluate the observation, and an additional appraiser is needed.

With more and more teachers advancing on the career ladder, some districts have found it necessary to adopt stricter TTAS performance criteria, where higher scores are required for placement and/or maintenance on the various levels, or to rank order teachers and fund awards for only a portion of teachers that qualify under state and district criteria. This increased competition for career ladder placements

contributes to the interest in appraiser reliability and scoring issues. Little information is known about the consistency of scores or differences between sub-groups of teachers on the Texas Teacher Appraisal System, particularly since the instrument has been revised to its current state.

As noted, there is a concern among teachers, administrators, teacher organizations, and policy makers about the reliability of observational data. Many feel that scores among appraisers should be similar for the same teacher. However, interrater reliability cannot be calculated because state law does not permit appraisers to view the same lesson. It is important to understand whether appraisers are evaluating similarly. The more similar a teacher's scores, the more "reliable" the assessment may appear. Because the TTAS scores are used for contract renewal, career ladder placements and maintenance, and staff development purposes, an accurate and reliable assessment of a teacher's performance is critical.

Purpose

The purpose of this study was to examine differences in Texas Teacher Appraisal System scores when considering the scores of the first four individual domains (Instructional Strategies, Management and Organization, Presentation of Subject Matter, and Learning Environment), the sum of the scores of Domains I through IV, and the overall summary performance score for teachers in a large school district over a period of two years. Specifically, scores between appraiser types (primary or second), levels (elementary or secondary), years (1988-1989 and 1989-1990) and appraisal periods (first or second) were investigated.

Method

In this study, differences in Texas Teacher Appraisal System scores of teacher subgroups over a period of two years were examined. Data were collected on each teacher in the district for the 1988-1989 and 1989-1990 school years and were analyzed according to scores of individual domains, sum of the scores of the first four domains, and overall summary performance score. Elementary and secondary teachers were analyzed separately in most cases.

Subjects

The subjects were the teachers of a large school district in central Texas for the school years of 1988-1989 (N = 1159) and 1989-1990 (N = 1207). All teachers were evaluated with the TTAS by a certified appraiser. In addition, employee information for each teacher was maintained. All subjects had existing data on school district computer files. The school district granted permission to allow the employee information and appraisal data to be used for research purposes.

Texas Teacher Appraisal System

The Texas Teacher Appraisal System (TTAS) was implemented statewide in 1986. Performance scores on the TTAS are used for Career Ladder placement and staff development purposes, and may be used for contract renewal. The instrument is categorized into five major domains which include the following:

Domain I -- Instructional Strategies

Domain II -- Classroom Management and Organization

Domain III -- Presentation of Subject Matter

Domain IV -- Learning Environment

Domain V -- Professional Growth and Responsibilities

Within each domain are broad criteria (a total of 13 criteria for the 5 domains) which are further broken down into behavioral indicators (a total of 65 indicators). When the instrument is scored, each indicator is worth one point and each criteria in Domains I through IV has three "exceptional quality" points that may be awarded at the criterion level. A total of 76 points is available in Domains I through IV, and a total of 92 points is possible in the entire instrument.

Domains I through IV are scored by both appraisers and reflect teacher behavior during the observation period. Domain V is scored only by the primary appraiser and involves job-related behaviors that are not directly observable in the lesson. An example of such a behavior is "Progresses in growth requirements or none needed."

Teachers on Levels 2 or 3 of the Texas Career Ladder are required to have two observations per year (typically one per semester), one by the primary appraiser (usually the teacher's supervisor) and one by the second appraiser (an assistant principal or off-campus administrator), as long as the teacher's performance is higher than satisfactory. Teachers on Level 1 of the career ladder are required to have four observations per year -- half of which are performed by the primary appraiser and half of which are performed by the second appraiser. One observation by each appraiser is conducted in each of the two appraisal periods, which parallel the fall and spring semesters. Although teachers of equal or higher career ladder levels are allowed to serve as appraisers for other teachers (State Board of Education Rule sec. 149.43, 1990), the district in this study utilizes only district administrators as appraisers.

In the school district studied, the appraisal process differed at the elementary level and the secondary level. At the elementary level, the primary appraiser was the campus principal, and the second appraiser was an off-campus administrator or district appraiser (the district employs two full-time appraisers). At the secondary level, the primary supervisor was the campus principal or assistant principal (grade level principal), and the second appraiser was generally another on-campus administrator (except in certain specialized areas, such as fine arts or special education).

Appraisers schedule and observe 45-minute lessons independently. Scoring is done on each observation, yet all scores of all observations are used to calculate the overall summary performance score for each teacher. This summative score is weighted, with 60% and 40% of the overall summary performance score accounted for by the primary appraiser and second appraiser, respectively.

Prior to certification as an appraiser, individuals completed approved training, passed a test on procedures, passed a scoring proficiency test on a videotaped lesson, and performed a field test with at least two other appraisers on an actual lesson. Yearly updates and recertification every three years are required for appraisers to maintain their certifications. All appraisers in this study were certified.

Appraisal Process

Teachers received TTAS orientations prior to observations being conducted. Teachers also received yearly updates on the procedures and TTAS instrument. Prior to each observation conducted during the school year, teachers participated in a pre-observation conference with the appraiser. Following the observation, the teacher and appraiser met for a post-conference to review the teacher's score on that

observation. Teachers could appeal the appraiser's score or procedure through the local grievance procedure. Information regarding the teacher's overall summary performance score and career ladder status was presented by the primary appraiser at the summative conference at the conclusion of the appraisal year. Primary appraisers were responsible for inputting all appraisal data into the district computer files.

Integrating Data

1988-1989 Data. Data from the school district for the school year 1988-1989 was contained on the Monthly Employee File and the Appraisal Record File for that school year. The Monthly Employee File was regularly maintained by the Personnel Department for record keeping on each school district employee. The existing data of interest in the Monthly Employee File included the following information for each teacher employed by the district: social security number, campus number, job code, ethnicity, district experience, total experience, date of birth, gender and career ladder level. All information for these variables was obtained from this file. No names were included on the file.

The Appraisal Record File contained the following information for each observation of each teacher: teacher social security number, appraiser social security number; whether the appraiser was the primary or second appraiser; which observation (first or second) by the appraiser; individual domain scores; and overall summary performance score. The information in the Appraisal Record File was input by primary appraisers after all observations of each teacher were completed.

Dates of each observation were not included in either file. After receiving advice from legal counsel for the district and permission was granted by the district superintendent, the principal investigator and the Director of Personnel Services

visited each of the 27 campuses. Each observation record and evaluation record for each teacher was viewed by the Director of Personnel Services to identify the date of the observation. The researcher did not have access to teacher or appraiser names; only the dates of the observations. The information was recorded on data sheets by the principal investigator. This data collection took place in December, 1989 through February, 1990.

1989-1990 Data. As a result of the 1988-1989 data collection, permission was granted by the district superintendent to revise the input screen for the Appraisal Record File so that the primary appraisers also would input the date of each observation. Thus, the visits to each campus to obtain the dates were not necessary for the second year data. The input into the Appraisal Record File was done by the primary appraisers in April and May, 1990.

Preparing Data for Analysis

The school district files were used to create appropriate computer files for data analysis. For some analyses (between appraisal periods, for example), separate files were created for teachers with four observations. Demographic data were combined with appraisal data. For all analyses, except in cases where the dependent variables were scores across years, data from 1988-1989 and 1989-1990 were combined. Additional data files were created for the various analyses.

Data Analysis

As Burstein (1980) has noted, much educational research occurs using multilevel data. In this quasi-experimental study, both multiple dependent and multiple independent measures were used. Where applicable, independent measures were combined for multiway analysis. Because some dependent variables were

linear combinations of others, data from this study were analyzed separately for the three sets of dependent variables: a) scores on the four individual domains (Instructional Strategies, Classroom Management and Organization, Presentation of Subject Matter, and Learning Environment); b) sum of scores of Domains I through IV; and c) overall summary performance score.

Due to the complexity of the analyses, multiple computer files were created from the existing files to combine years (1988-1989 and 1989-1990) and to isolate teachers with two or four observations at each level (elementary or secondary). In addition, files were created where information was matched across years for the various subgroups of teachers. In cases where a subject was missing information, the record was deleted from the analysis. Elementary and secondary teachers were analyzed separately in all instances, as were teachers with two observations and teachers with four observations (except when the overall summary performance score was examined). In instances where teachers with two or four observations were combined to make comparisons across years, subjects were not included if they received four observations the first year and two observations the second year. All repeated measures MANOVAs were computed using a mixed multivariate model (Schutz & Gessaroli, 1987).

Differences in scores between appraiser types and levels. For the scores of the individual domains, a 2X2 (level: elementary/secondary X appraiser type: primary/second) MANOVA with repeated measures on the second factor was computed. For the sum of Domains I through IV, a 2X2 (level: elementary/secondary X appraiser type: primary/second) ANOVA with repeated measures on the second factor was computed. ANOVA also was computed for the

overall summary performance score. Follow-up to MANOVA was conducted as specified by Bray and Maxwell (1982). Significant between factors in the multiway MANOVAs were followed by stepwise discriminant function analysis. ANOVA was performed for the first dependent variable that entered the equation. Then, for all other significant predictors, ANCOVA was used with the variables that entered the equation previously as covariates. For significant within factors in MANOVA, repeated measures ANOVA was used to determine for which variables significance existed. For all analyses, alpha was set at .01.

Differences in scores between appraisal periods and years. To determine if there were differences between appraisal periods, only teachers with four observations were used. By state mandate, teachers with two observations are not observed across appraisal periods by the same appraiser. Data were combined across years. MANOVA with repeated measures (alpha = .01) was used to determine if there were differences for the scores on the four individual domains. Repeated measures ANOVA was used to follow-up the significant MANOVA. Separate repeated measures ANOVAs were used to determine differences between appraisal periods for the sum of the four domains and the overall summary performance score.

Similar analyses were used to examine differences across years: (a) MANOVA with repeated measures for the scores on the four domains; (b) repeated measures ANOVA to follow-up the significant MANOVA; (c) repeated measures ANOVA for the sum of the four domains and the overall summary performance score.

Results

In cases where it was more appropriate to use least squares means in the analyses (in follow-up ANCOVAs), significant differences are reported according to the least squares means. Means are reported in the appropriate tables.

Descriptive Information

For 1988-1989, teachers had a mean overall summary performance score of 166.39 (out of 184) and a mean of 169.72 for 1989-1990. Table 1 summarizes the overall descriptive statistics, including means and standard deviations, for the three sets of dependent variables (scores on the individual domains, sum of Domains I through IV, and overall summary performance score) by year, and for the combined years (1988-1989 and 1989-1990) included in the study.

Insert Table 1 about here

Differences in Scores Between Appraiser Types (Primary or Second)

Scores of the first four domains -- Instructional Strategies, Management and Organization, Presentation of Subject Matter, Learning Environment. For this dependent variable, MANOVA was computed for the various subgroups of teachers. Appraiser type was a repeated measure in these analyses. When elementary teachers ($N = 771$) and secondary teachers ($N = 763$) with two observations were examined, a significant difference ($F[4,1529] = 23.30$, $p < .01$) between appraiser type was found. Follow-up analyses indicated this difference was explained by Domain I (Instructional Strategies) and Domain II (Management and Organization). There were no significant differences in the other two domains. Elementary teachers with

two observations received higher scores from the primary appraiser in all domains, although significance was found for Domains I and II only. Secondary teachers received significantly higher scores from the second appraiser in Domain I. Teachers received higher scores from the primary appraiser in all other domains, with Domain II being significant. Table 2 includes the means and standard deviations of each domain by level (elementary or secondary) and appraiser type (primary or second) for teachers with two observations.

Insert Table 2 about here

When elementary teachers (N = 960) and secondary teachers (N = 704) with four observations were examined, MANOVA also indicated a significant difference (F[4, 827] = 6.18, p<.01) between appraiser type. As with the teachers with two observations, the difference was explained by Domain II. Elementary and secondary teachers received higher scores from the primary appraiser in all domains, except Domain I, where the second appraiser scores were slightly higher, but these were not significant. Table 3 includes means and standard deviations of each domain by level and appraiser type for teachers with four observations.

Insert Table 3 about here

Sum of Domains I through IV. ANOVA identified a significant difference (F[1, 1532] = 51.44, p<.01) in the scores awarded by primary and second appraisers for elementary teachers (N = 771) and secondary teachers (N = 763) with

two observations . Primary appraiser scores were significantly higher than those of the second appraisers. Table 4 includes the means and standard deviations of the sum of Domains I through IV by level and appraiser type.

Insert Table 4 about here

ANOVA also indicated a significant difference ($F[1, 830] = 8.04$, $p<.01$) in scores awarded by primary and second appraisers for elementary teachers ($N = 960$) and secondary teachers ($N = 704$) with four observations. Again, primary appraisers awarded higher scores, as presented in Table 5.

Insert Table 5 about here

Differences in Scores Between Levels (Elementary or Secondary)

Scores of the first four domains -- Instructional Strategies, Management and Organization, Presentation of Subject Matter, Learning Environment. A significant difference($F[4, 1529] = 10.44$, $p<.01$). in level (elementary or secondary) was demonstrated through MANOVA for teachers with two observations. Stepwise discriminant function analysis (significance level to enter = .01) identified Domain I, Domain II and Domain IV, respectively, as the best predictors of level. ANOVA indicated a significant difference ($F[1, 3066] = 28.03$, $p<.01$) in level in Domain I, with elementary teachers receiving higher scores. ANCOVA was performed on the next two domains to enter. Domain II showed a significant difference ($F[1, 3065] = 6.99$, $p<.01$) in level, with secondary teachers receiving higher scores on this

domain. Domain IV was the last domain to enter, and ANCOVA indicated a significant difference ($F[1, 3064] = 11.19, p<.01$) in levels, with elementary teachers receiving higher scores. Table 6 includes means and standard deviations for Domain I, II, III and IV by level.

Insert Table 6 about here

For teachers with four observations, MANOVA demonstrated a significant difference ($F[4, 827] = 5.16, p<.01$) in levels. Stepwise discriminant function analysis identified Domain I was the first and only domain to enter. ANOVA on Domain I showed a significant difference ($F[1,3326] = 20.56, p<.01$) in the elementary and secondary scores, with elementary teachers receiving higher scores. Table 7 includes means and standard deviations for Domain I, II, III and IV by level.

Insert Table 7 about here

Sum of Domains I through IV. ANOVA for teachers with two observations indicated a significant difference ($F[1, 1286] = 12.33, p<.01$) between elementary teachers ($N = 1264$) and secondary teachers ($N = 1312$) for this dependent variable. Elementary teachers received higher scores. For teachers with four observations, ANOVA again showed a significant difference ($F[1, 942] = 6.92, p<.01$) in levels. Elementary teachers again ($N = 1064$) had significantly higher scores than secondary teachers ($N = 824$). Means and standard deviations for elementary teachers and secondary teachers with two and four observations are included in Table 8.

Insert Table 8 about here

Overall summary performance score. ANOVA demonstrated a significant difference ($F[1, 878] = 7.53$, $p<.01$) across levels for this dependent variable. The mean for elementary teachers was 170.1421 and the mean for secondary teachers was 168.37495. The highest overall summary performance score a teacher can receive is 184.

Differences in Scores Between Appraisal Periods (First or Second)

The analyses for this hypothesis, with the two sets of dependent variables (scores of the first four domains and sum of Domains I through IV), included only those teachers with four observations. Teachers with two observations receive one observation in each appraisal period by different appraisers (primary and second). Those teachers with four observations receive one observation by the primary appraiser and one observation by the second appraiser in each of the two appraisal periods. Therefore, the two appraisal periods may be compared.

Scores of the first four domains -- Instructional Strategies, Management and Organization, Presentation of Subject Matter, Learning Environment. MANOVA, with appraisal period as a repeated measure, on the first four domains showed there were significant differences ($F[4, 827] = 49.55$, $p<.01$) between appraisal periods (first or second) for all teachers with four observations in all domains. For elementary teachers ($N = 960$) and for secondary teachers ($N = 704$), the scores received for each of the four domains were significantly higher ($p<.01$) in the second

appraisal period. Domain means and standard deviations by level and appraisal period are summarized in Table 9.

Insert Table 9 about here

Sum of Domains I through IV. ANOVA for elementary teachers indicated a significant difference ($F[9, 3239] = 4.25$, $p<.01$) between the two appraisal periods. A significant difference ($F[9, 2663] = 3.19$, $p<.01$) for secondary teachers was seen as well. In both cases, the second appraisal period produced higher scores. Table 10 includes the means and standard deviations by level and appraisal period for this dependent variable.

Insert Table 10 about here

Differences in Scores Between Years (1988-1989 or 1989-1990)

Sum of Domains I through IV. For elementary teachers ($N = 632$) and secondary teachers ($N = 656$) with two observations, ANOVA showed a significant difference ($F[1, 1286] = 109.35$, $p<.01$) between the two school years (1988-1989 and 1989-1990) examined in this study. In both cases, the scores received in the second year were higher. Means and standard deviations by level and year are included in Table 11 for teachers with two observations. A similar result was seen with elementary teachers (N = 532) and secondary teachers (N = 412) who received four observations, with significant differences ($F[1, 942] = 2.87$, $p<.01$) between years. Again, the scores received in the second year were higher. Means and

standard deviations by level and year are included in Table 12 for teachers with four observations.

Insert Tables 11 and 12 about here

Overall summary performance score. ANOVA indicated a significant difference ($F[1, 878] = 245.01$, $p<.01$) between the two school years for elementary ($N = 449$) and secondary ($N = 431$) teachers. Overall summary performance scores for the second year (1989-1990) were higher. Means and standard deviations by level and year are presented in Table 13.

Insert Table 13 about here

Discussion

It should be noted that many significant differences were seen. Overall, the scores were extremely high and the significant differences involved less than one point when domain scores or the sum of the first four domains were examined. Each indicator in the instrument is worth one point, and each exceptional quality award is worth three points. Thus, the significant differences found often were less than the difference between receiving exceptional quality credit on one criteria, or not receiving the exceptional quality credit.

Differences in Scores Between Appraiser Type (Primary or Second)

Previous research on the Texas Teacher Appraisal System indicated that primary appraisers tended to award higher scores (Holmes, 1989; Performance

Assessment Systems, Inc., 1987). This is consistent with literature on other assessment instruments as well, where the on-campus evaluator's or principal's scores were higher (Cronin & Capie, 1986; Ellett & Capie, 1985; Kelley, 1985; Rose & Huynh, 1984; Wise et al., 1984). The results of this study paralleled this research. For the most part, when scores of domains and the sum of the scores of the first four domains were examined, the primary appraiser awarded higher scores.

In the case of the TTAS, part of the explanation may be attributed to "walk-throughs" performed by the primary appraiser. By state law, second appraisers score only the 45 minute observation; whereas primary appraisers may change the teacher's score as a result of walk-throughs that may occur throughout the appraisal period (Texas Education Code, 1990). It is thought by teachers and appraisers that the purpose of the walk-through is to attempt to observe behaviors that were not apparent during the formal observation. In that case, the teacher's score would be raised as a result of a subsequent walk-through. However, the walk-through may raise or lower the teacher's score. It is generally thought that the score is raised. Consequently, the primary appraiser's score may be higher, particularly if the primary and second appraiser awarded similar scores from the 45 minute formal observation.

It is also thought that the primary appraiser, or on-campus evaluator, may be more familiar with the teacher's overall abilities and may actually score the teacher on overall performance, rather than the formal observation (Holmes, 1989). The second appraisers, however, have no basis for awarding scores other than the formal observation. If the primary appraiser can more accurately judge a teacher's performance because of familiarity, it would be assumed that the primary appraiser's

score would be more valid. Doyle (1983) suggests that the greater the experience the evaluator has with the teacher, the more dependable the evaluation. Some feel that assessments performed by evaluators from other buildings or districts are more objective, and teachers' morale is higher throughout the assessment process (Association of Teacher Educators, 1988).

Literature also suggests that role conflict may exist for principals or primary appraisers (Duckett, 1985; Stanley & Popham, 1988). The principal is the "helper person," who serves as the instructional leader of the campus, and who also may fulfill many other roles in a helping capacity, such as supporter, counselor, etc. for the teacher. The principal is also the "hatchet person," who makes decisions regarding whether to rehire, non-renew or terminate the teacher. This role conflict may result in the principal awarding higher scores, in that the on-campus principal must interact with the teacher on a daily basis. The second appraiser, however, spends little time on the campus and interacts with each teacher infrequently. This lack of face-to-face contact may enable the second appraiser to score a teacher objectively, without anticipating consequences of a low score on the relationship with the teacher. After implementation of the TTAS, principals expressed concern that the greater formality of the instrument hampered relationships with teachers (Bezdek, Cross & Guerrero, 1988).

MANOVA indicated a significant difference between the two appraisers' scores when scores of the individual domains were examined. For teachers with two observations, the difference was explained by the first two domains, Instructional Strategies and Management and Organization. Teachers with four observations also received significantly higher scores from the primary appraiser, explained by Domain

II. Within this domain is Criterion 5 (Manages Student Behavior) that includes indicators or teacher behaviors that involve discipline strategies. The exceptional quality credit associated with this criterion has been referred to as the "super EQ," in that it is very difficult for the second appraiser to award exceptional quality credit for this criterion if there is no off-task or inappropriate behavior, and the teacher has no need to redirect students or reinforce desired behavior (Crain, 1990). If the appraiser awards exceptional quality credit in all or most other criteria, then the appraiser may feel confident in awarding the additional credit on Criterion 5 as well. If exceptional quality credit is awarded for any criterion, the appraiser must document on the evaluation or observation record the justification for the award. If no inappropriate student behavior occurs, it may be more difficult to justify and document the rationale for the award. The primary appraiser is on campus on a daily basis and is much more familiar with the teacher's discipline strategies and management plan and may be more apt to award exceptional quality credit as they frequently observe the teacher handling discipline matters. In addition, if off-task or inappropriate behavior occurs during the observation, the primary appraiser may be more lenient in scoring if he or she is aware of a possible cause of the student behavior, such as a fire drill, assembly, or emergency of which the second appraiser may not be aware.

The results demonstrated that secondary teachers with two observations received higher scores from the second appraisers in Domain I. Although this difference between the two appraisers' scores was .1651 points, the difference was significant. When teachers with four observations were examined, the second appraisers again awarded just slightly higher scores for Domain I, although this difference was not significant. This is very difficult to explain. The primary

appraiser's familiarity with the teacher may work to the teacher's disadvantage in this circumstance. Domain I, Instructional Strategies, includes indicators that involve teaching behaviors such as monitoring, interacting, communicating learning expectations, wait time, etc. Many times, teachers make a much greater effort to "get out from behind the desk," to interact while students are completing assignments, or to wait the standard 3 to 5 seconds for a student response while the appraiser is in the classroom. Appraisers who see the teacher on a regular basis may know that these behaviors do not occur all the time, and the appraisers may be more hesitant to award exceptional quality credit. Second appraisers, who do not know the teacher does not do these things everyday, are more likely to award exceptional quality credit if the behaviors are observed. When all domains are considered, however, the primary appraiser awards significantly higher scores.

Differences in Scores Between Levels (Elementary or Secondary)

In the current study, elementary teachers received higher scores in almost every case. When other observation instruments have been examined, elementary teachers also tend to receive higher scores than secondary teachers (Alabama Career Incentive Program Project Staff, 1987; Dickson & Wiersma, 1984; Florida Coalition for the Development of a Performance Measurement System, 1984; Ligon & Ellis, 1986; Tadlock & Nesbit, 1984). Pilot studies on the TTAS by Performance Assessment Systems, Inc. (1986) and research on the TTAS prior to its revision in 1987 also demonstrated higher scores for elementary teachers (Holmes, 1989).

When individual domains were examined with MANOVA for teachers with two observations, a significant difference between levels was seen, and Domains I, II and IV explained this difference. In Domain I and Domain IV, elementary teachers

received significantly higher scores. Secondary teachers received higher scores in Domain II. Although the difference was not significant for Domain III, elementary teachers again received higher scores. Elementary teachers with four observations scored higher in all domains, except for Domain III. The difference was significant for Domain I only, however. In both remaining sets of dependent variables (sum of Domains I through IV and overall summary performance score) and in both subgroups of teachers (those with two observations and those with four observations), elementary teachers received significantly higher scores.

Much of the research base used in the development of many observation instruments involved elementary level instruction (Holdzkom, 1991). It could be that many indicators included in observation instruments are behaviors that are typically seen in effective elementary classrooms. Consequently, secondary teachers, who do not typically display some of those behaviors, might not score as well. Research suggests that different teaching behaviors are more and less effective for students at different grade levels (Brophy & Good, 1986). Because of the greater amount of research conducted at the elementary level and the possible differences in effective teaching behaviors at the secondary level, it could be that secondary teachers are demonstrating behaviors that are appropriate, but not included on the TTAS.

Some behaviors on the part of the students may also differ at the secondary level. The literature suggests that older students need less overt participation (Brophy & Good, 1986). This more subtle student involvement may be more difficult for the appraiser to detect and subsequently score. The cause for lower scores by secondary teachers could be that they, indeed, display fewer of the

effective teaching behaviors on the TTAS, or less overt happenings may go unnoticed.

Elementary teachers are responsible for delivering instruction for a variety of subject areas. In the district under study, teachers generally had a choice for the formal observation time period. Even if the choice was not provided, elementary teachers could alter the daily schedule so that they could select the subject area to teach. This greater flexibility allows elementary teachers to pick a subject area in which they may feel more confident or which may promote student enjoyment, participation, and visible success. Secondary teachers, however, often teach within one subject area and may seem limited in choices for lessons.

Secondary teachers with two observations did receive significantly higher scores in Domain II than their elementary counterparts. As noted, this domain contains disciplinary strategies and management behaviors of the teacher. Typically, secondary teachers select their "best" class for the formal observation, whereas elementary classroom teachers have the same students all day. Secondary teachers are likely to select the class which have the best behaved and most attentive students.

Differences in Scores Between Appraisal Periods (First or Second)

This analysis involved only teachers with four observations so that primary and second appraiser scores should be compared between the two appraisal periods. Thus, for the most part, teachers at Level 1 and 0 on the career ladder were involved; although, teachers on Levels 2 and 3 were included if they had four observations.

The TTAS was implemented not only as an evaluation tool, but for instructional improvement as well. Teachers who score below a pre-set criteria are required to have a professional growth plan for individual improvement (Texas

Education Code sec. 13.303, 1990). In addition, campus and district staff development may be initiated for components or areas of the TTAS to facilitate individual and group growth.

Despite the intent of the instrument to help teachers improve instructionally, there continues to be concern when a teacher's scores are not consistent during the year. When an appraiser's scores differ from semester to semester, teachers often assume that the lowest score was incorrect, particularly if the teacher received a higher score the previous year. However, when the appraiser's score is higher the second appraisal period, it is far more palatable to the teacher than when the reverse occurs. The assumption by teachers (and some appraisers) is that teachers may improve, but can never "get worse." Therefore, there exists a concern for consistency of scores and a concern for instructional improvement.

When scores (elementary and secondary) on the individual domains and the sum of the domains were examined, the second appraisal period scores were significantly higher. Teachers' scores significantly increase during the year, suggesting instructional improvement. However, instructional improvement may be interpreted in several ways. Teachers may be performing better for the appraiser, particularly after the first observation and the appraiser has conferenced and perhaps suggested ways to improve. Teachers may become more familiar with the instrument, particularly since the teachers are mainly Level 1 teachers who generally have fewer years experience with the instrument. Appraisers may feel obligated to award a higher score as the year progresses because teachers can and do improve, and the appraiser has conferenced and perhaps worked individually with the teacher. In addition, the appraiser may avoid the conflict that could arise if a teacher's score is

lowered the second semester. As discussed earlier, role conflict of the appraiser being the "helper person" and the "hatchet person" is a concern (Duckett, 1985; Stanley & Popham, 1988).

Research on stability of behavior demonstrates that teachers differ on different occasions (Cronin & Capie, 1986). It is also suggested that variation in teacher behavior should be expected, as teachers use different instructional arrangements, particularly at the elementary level where different subjects are taught (Stodolski, 1984). The current research was consistent. Scores were significantly different between the two appraisal periods, with the second appraisal period producing the higher scores.

Differences in Scores Between Years (1988-1989 or 1989-1990)

For this analysis, teachers were omitted if they were not employed by the district for both school years (1988-1989 and 1989-1990) studied. Thus, analyses across years compared teachers' scores from year to year.

For both levels, and for both sets of dependent variables (sum of Domains I through IV, overall summary performance score), scores for the second year were significantly higher than scores for the first years. Thus, not only do the scores increase significantly from the first appraisal period to the second, scores increase over years. The increase could, again, be due to true instructional improvement, expectations of the appraiser, or the teacher's greater familiarity with the instrument. As will be discussed later, the continual increase in scores is very problematic for districts who have limited funding for career ladder supplements.

Instructional improvement could occur through a variety of methods -- conferencing with the appraiser, greater familiarity with the expectations of the

instrument, staff development, or through observations of other teachers. Appraisers may expect teachers to improve due to the efforts of the appraiser and may award higher scores through the years. Teachers who show no improvement over a period of time sometimes do not continue at the same campus or in the profession.

Implications

Although the purposes of teacher evaluation are to improve the quality of teaching and serve as the basis for administrative decisions, the assessment may become meaningless if scores continue to increase and consistent subgroup differences occur. Because its results are tied to career ladder and financial bonuses in Texas, teacher evaluation is an issue of importance and controversy. The results of this study suggest that statewide teacher assessment scores are problematic.

Not only do teachers' scores increase during the year, from the first appraisal period to the second, they increase from year to year. This continual and significant increase in scores makes administrative decisions more difficult. As scores increase and more of a ceiling effect is seen, scores may lose practical meaning because little differentiation among teachers will exist. Competition for career ladder supplements will increase, as districts will find funding more and more limited with more teachers qualifying for advanced levels of the career ladder. The power to make contract renewal decisions based on teaching performance also will be limited. If this trend continues, instrument validity may be questioned. The results of this study suggest a modification of the instrument or a modification in scoring so that scores are meaningful, and improvement of teaching performance is reflected.

In addition, the results demonstrate subgroup differences in scores received on the TTAS, consistent with results of other evaluation instruments. Differences

between appraiser type should be addressed through training of appraiser at the state and local level. Differences in scores received by elementary and secondary teachers suggest further study on the validity of the instrument across grade levels, training of appraisers, and possible staff development for teachers. Additional research on teacher evaluation and the Texas Teacher Appraisal System can only benefit policy-makers and those implementing large scale evaluation systems.

References

Alabama Career Incentive Program Project Staff. (1987). Alabama performance-based career incentive program: A report on the norming study of the spring semester. Atlanta, GA: Georgia State University. (ERIC Document Reproduction Service No. ED 302 578)

Association of Teacher Educators. (1988). Teacher assessment. Reston, VA: Author. (ERIC Document Reproduction Service No. ED 289 869)

Bezdek, R., Cross, R., & Guerrero, T. (1988, April). Bureaucratic control and principal role. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. ED 303 897)

Bray, J. H., & Maxwell, S. E. (1982). Analyzing and interpreting significant MANOVAs. Review of Educational Research, 52, 340-367.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. C. Wittrock (Ed.), Handbook of research on teaching (3rd ed.). New York: Macmillan.

Burstein, L. (1980). The analysis of multilevel data in educational research and evaluation. Review of Research in Education, 8, 158-233.

Crain, J. (1990). Texas Orientation 1990 Texas Teacher Appraisal System [Videotape]. Austin, TX: Texas Education Agency.

Cronin, L. L., & Capie, W. (1986, April). The influence of daily variation in teacher performance on the reliability and validity of assessment data. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 274 704)

Dickson, G. E., & Wiersma, W. (1984). Empirical measurement of teacher performance. Research and evaluation in teacher education. Toledo, OH: Toledo University, Center for Educational Research and Services. (ERIC Document Reproduction Service No. ED 263 211)

Doyle, K. (1983). Evaluating teaching. Toronto: D. C. Heath.

Duckett, W. R. (Ed.). (1985). The competent evaluator of teaching: A CEDR monograph. (Report No. ISBN-0-87367-720-X). Bloomington, IN: Phi Delta Kappa, Center on Evaluation and Research. (ERIC Document Reproduction Service No. ED 266 536)

Ellett, C. (1990). A new generation of classroom-based assessments of teaching and learning: Concepts, issues and controversies from pilots of the Louisiana STAR. Baton Rouge, LA: Author.

Ellett, C. D., & Capie, W. (1985, April). Assessing meritorious teacher performance: A differential validity study. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 264 255)

Ellett, C. D., & Garland, J. S. (1986). Teacher evaluation practices in our largest school districts: Are they measuring up to "state-of-the-art" systems? Baton Rouge, LA: Author. (ERIC Document Reproduction Service No. ED 294 316)

Florida Coalition for the Development of a Performance Measurement System. (1984). Teacher evaluation study. Norming study, generic competence revision, feedback proposal, cognitive examinations, and specialized domains. Report for 1983-1984. Tallahassee, FL: Author. (ERIC Document Reproduction Service No. ED 266 122)

Holdzkom, D. (1991). Teacher performance appraisal in North Carolina: Preferences and practices. Phi Delta Kappan, 72, 782-785.

Holmes, G. E. (1989). The effects of appraiser, contextual, and procedural variables on the performance rating of teachers as determined by the Texas Teacher Appraisal System . Dissertation Abstracts International, 49, 2470A-2471A. (University Microfilms No. 8819501)

Kelley, M. F. (1985, April). The Arizona performance-based teacher certification program. Paper presented at the Annual Meeting of the American Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED 263 078)

Ligon, G., & Ellis, J. (1986, April). Adjusting for rater bias in teacher evaluations: Political and technical realities. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco. (ERIC Document Reproduction Service No. ED 273 677)

Medley, D. M., & Mitzel, H. E. (1963). Measuring classroom behavior by systematic observation. In N. L. Gage (Ed.), Handbook of research on teaching (pp. 247-328). Chicago: Rand McNally.

Performance Assessment Systems, Inc. (1986). A summary analysis of the Texas Teacher Appraisal System (TTAS): Observations in six pilot districts during 1985. Athens, GA: Author.

Performance Assessment Systems, Inc. (1987). An interpretive summary of 1986-1987 TTAS data. Athens, GA: Author.

Rose, J. S., & Huynh, H. (1984, April). Technical issues for adopting the APT for districtwide teacher evaluation. Paper presented at the Annual Meeting of the National Council on Measurement in Education, New Orleans. (ERIC Document Reproduction Service No. ED 247 266)

Rosenshine, R., & Furst, N. (1973). The use of direct observation to study teaching. In M. W. Travers (Ed.), Second handbook of research on teaching (pp. 122-183). Chicago: Rand McNally.

Schutz, R. W., & Gessaroli, M. E. (1987). The analysis of repeated measures designs involving multiple dependent variables. Research Quarterly for Exercise and Sport, 58, 132-149.

Stanley, J. J., & Popham, J. W. (Eds.). (1988). Teacher evaluation: Six prescriptions for success. Alexandria, VA: Association for Supervision and Curriculum Development. (ERIC Document Reproduction Service No. ED 299 683)

State Board of Education Rule, sec. 149.41-149.44 (1990).

Stodolski, S. S. (1984). Teacher evaluation: The limits of looking. Educational Researcher, 13(9), 11-18.

Tadlock, J., & Nesbit, L. (1984, November). The relationship of teacher evaluation scores generated by a process-product evaluation instrument to selected variables. Paper presented at the Annual Meeting of the Mid-South Educational Research Association, New Orleans. (ERIC Document Reproduction Service No. 252 931)

Texas Education Agency. (1988). Teacher orientation manual 1988-1989. Austin, TX: Author.

Texas Education Code, sec. 13.302-13.321 (1990).

The University of Texas at Austin Computation Center. (1989). MS-DOS Kermit pocket reference list, CCRL-36. Austin, TX: Author.

United States Department of Education. (1987). What's happening in teacher testing. An analysis of state teacher testing practices. Washington, DC: Author.

Wise, A. E., Darling-Hammond, L., McLaughlin, M. W., & Bernstein, H. T. (1984). Teacher evaluation: A study of effective practices. Santa Monica, CA: Rand Corporation. (ERIC Document Reproduction Service No. ED 246 559)

Table 1.

Overall Descriptive Statistics

By Dependent Variable and Year(s)

Means and Standard Deviations of Domains I Through IV	Mean	Standard Deviation
1988-1989 (N = 3148)		
Domain I	16.2122	2.1875
Domain II	20.5511	2.9013
Domain III	17.3094	2.2900
Domain IV	12.4041	1.9651
1989-1990 (N = 3248)		
Domain I	16.6438	1.9704
Domain II	21.0219	2.8139
Domain III	17.8347	2.1763
Domain IV	12.7608	1.7645
Combined Years (N = 6396)		
Domain I	16.4314	2.0911
Domain II	20.7902	2.8667
Domain III	17.5761	2.2482
Domain IV	12.5852	1.8743

Table 1.

(con't)

Means and Standard Deviations of the Sum of Domains I Through IV

	N	Mean	Standard Deviation
1988-1989	3148	66.4768	6.9948
1989-1990	3248	68.2611	6.4440
Combined Years	6396	67.3829	6.7791

Means and Standard Deviations of Overall Summary Performance Score

	N	Mean	Standard Deviation
1988-1989	1159	166.3941	11.3969
1989-1990	1207	169.7223	10.6002
Combined Years	2366	168.0920	11.1206

Table 2.

Means and Standard Deviations of Domains I Through IV by Level and Appraiser Type

Teachers With Two Observations

	Mean			
	Domain I	Domain II	Domain III	Domain IV
Elementary (N = 771)				
Primary	17.1492* (1.5782)	21.9144* (2.5263)	18.2335 (1.9811)	13.1284 (1.5282)
Second	16.8016* (1.8272)	20.8872* (2.5393)	17.9546 (2.0153)	12.9896 (1.5708)
Secondary (N = 763)				
Primary	16.5400* (2.0110)	21.7274* (2.4637)	17.9987 (2.1829)	12.8453 (1.7598)
Second	16.7051* (1.9226)	21.2254* (2.6937)	17.9659 (2.1569)	12.7300 (1.8042)
Points Available	18.0000	24.0000	20.0000	14.0000

*Significant difference (p<.01) between primary and second appraisers

Note: Values are means with standard deviations

Table 3.

Means and Standard Deviations of Domains I Through IV by Level and Appraiser Type

Teachers With Four Observations

	Mean			
	Domain I	Domain II	Domain III	Domain IV
Elementary (N = 960)				
Primary	16.2146 (2.2339)	20.5959* (3.0139)	17.1896 (2.3917)	12.4104 (1.9327)
Second	16.2688 (2.0899)	19.8792* (2.7423)	17.1032 (2.1417)	12.3459 (1.8525)
Secondary (N = 704)				
Primary	15.8736 (2.2990)	20.2088* (3.0729)	17.1918 (2.4364)	12.2244 (2.0742)
Second	15.9006 (2.3338)	20.0540* (3.0895)	17.1179 (2.2734)	12.0341 (2.1392)
Points Available	18.0000	24.0000	20.0000	14.0000

*Significant difference (p<.01) between the primary and second appraiser

Note: Values are means with standard deviations

Table 4.

Means and Standard Deviations of the Sum of Domains I Through IV by Level and Appraiser Type

Teachers With Two Observations

	Mean	Std Dev
Elementary (N = 771)		
Primary	70.4254*	5.6880
Second	68.6329*	5.7356
Secondary (N = 763)		
Primary	69.1114*	5.8802
Second	68.6265*	6.0391
Points Available	76.0000	

*Significant difference (p<.01) between the primary and second appraiser

Table 5.

Means and Standard Deviations of the Sum of Domains I Through IV by Level and Appraiser Type

Teachers With Four Observations

	Mean	Std Dev
Elementary (N = 960)		
Primary	66.4104*	7.4132
Second	65.5969*	6.5108
Secondary (N = 704)		
Primary	65.4985*	7.1941
Second	65.1065*	7.3379
Points Available	76.0000	

*Significant difference (p<.01) between the primary and second appraiser

Table 6.

Means and Standard Deviations of Domains I Through IV by Level and Domain

Teachers With Two Observations

	Mean			
	Domain I	Domain II	Domain III	Domain IV
Elementary (N = 1542)				
	16.9754*	21.4008*	18.0941	13.0590*
	(1.7155)	(2.5836)	(2.0025)	(1.5507)
Secondary (N = 1526)				
	16.6225*	21.4764*	17.9823	12.7877*
	(1.9684)	(2.5926)	(2.1693)	(1.7825)
Points Available	18.0000	24.0000	20.0000	14.0000

*Significant difference (p<.01) between the elementary and secondary level

Note: Values are means with standard deviations

Table 7.

Means and Standard Deviations of Domains I Through IV by Level and Domain

Teachers With Four Observations

	Mean			
	Domain I	Domain II	Domain III	Domain IV
Elementary (N = 1920)				
	16.2417*	20.2375	17.1464	12.3781
	(2.1627)	(2.9028)	(2.2699)	(1.8928)
Secondary (N = 1408)				
	15.8871*	20.1314	17.1548	12.1293
	(2.3157)	(3.0811)	(2.3558)	(2.1084)
Points Available	18.0000	24.0000	20.0000	14.0000

*Significant difference (p<.01) between the elementary and secondary level

Note: Values are means with standard deviations

Table 8.

Means and Standard Deviations of the Sum of Domains I Through IV by Number of Observations and Level

	N	Mean	Std Dev
Teachers With Two Observations			
Elementary	1264	70.0934*	5.3419
Secondary	1312	69.1662*	5.6518
Teachers With Four Observations			
Elementary	1064	66.0902*	6.8605
Secondary	824	65.0377*	7.6180
Points Available		76.0000	

*Significant difference ($p<.01$) between the elementary and secondary level

Table 9.

Means and Standard Deviations of Domains I Through IV by Level and Appraisal Period

Teachers With Four Observations

	Mean			
	Domain I	Domain II	Domain III	Domain IV
Elementary (N = 960)				
First	15.9521* (2.2511)	19.9063* (2.9411)	16.8021* (2.2830)	12.1886* (1.9453)
Second	16.5313* (2.0308)	20.5688* (2.8269)	17.4907* (2.2048)	12.5677* (1.8203)
Secondary (N = 704)				
First	15.6463* (2.3743)	19.7699* (3.1322)	16.9787* (2.3832)	11.9503* (2.1439)
Second	16.1278* (2.2314)	20.4929* (2.9879)	17.3309* (2.3163)	12.3082* (2.0582)
Points Available	18.0000	24.0000	20.0000	14.0000

*Significant difference (p<.01) between the first and second appraisal period

Note: Values are means with standard deviations

Table 10.

Means and Standard Deviations of the Sum of Domains I Through IV by Level and Appraisal Period

Teachers With Four Observations

	Mean	Std Dev
Elementary (N = 960)		
First	64.8490*	7.1010
Second	67.1583*	6.6771
Secondary (N = 704)		
First	64.3452*	7.4222
Second	66.2599*	6.9823
Points Available	76.0000	

*Significant difference (p<.01) between the first and second appraisal period

Table 11.

Means and Standard Deviations of the Sum of Domains I Through IV by Level and Year

Teachers With Two Observations

	Mean	Std Dev
Elementary (N = 632)		
1988-1989	69.3544*	5.6159
1989-1990	70.8323*	4.9483
Secondary (N = 656)		
1988-1989	68.3384*	5.5790
1989-1990	69.9939*	5.6069
Points Available	76.0000	

*Significant difference (p<.01) between 1988-1989 and 1989-1990

Table 12.

Means and Standard Deviations of the Sum of Domains I Through IV by Level and Year

Teachers With Four Observations

	Mean	Std Dev
Elementary (N = 532)		
1988-1989	64.4041*	7.1611
1989-1990	67.7763*	6.1028
Secondary (N = 412)		
1988-1989	62.9782*	7.7677
1989-1990	67.0971*	6.8824
Points Available	76.0000	

*Significant difference (p<.01) between 1988-1989 and 1989-1990

Table 13.

Means and Standard Deviations of Overall Summary Performance Score

by Level and Year

	Mean	Std Dev
Elementary (N = 449)		
1988-1989	168.1684*	10.6312
1989-1990	172.1158*	8.7918
Secondary (N = 431)		
1988-1989	166.1972*	11.2293
1989-1990	170.5527*	10.5427
Points Available	184.0000	

*Significant difference (p<.01) between 1988-1989 and 1989-1990